

1. Dolphin Scheduler

- 1.1. DS是什么?
- 1.2. 和其他调度系统对比
- 1.3. 为什么要用DS呢?
- 1.4. 谁在用DS呢?
- 1.5. 初次见面
- 1.6. 试用

1. Dolphin Scheduler

1.1. DS是什么?

A distributed and easy-to-expand visual DAG workflow scheduling system

Dedicated to solving the complex dependencies in data processing, making the scheduling system out of the box for data processing.

Apache DolphinScheduler 是一个分布式去中心化，中国人易观开源的一个分布式易扩展的可视化 DAG 工作流任务调度系统。致力于解决数据处理流程中错综复杂的依赖关系，使调度系统在数据处理流程中开箱即用。DolphinScheduler 曾用名为 "EasyScheduler"，由易观开发，美国时间 2019 年 8 月 29 日，正式通过顶级开源组织 Apache 基金会的投票决议，以全票通过的优秀表现正式成为 Apache 孵化器项目。由于名称已在国外某 App 使用，经社区讨论与投票后改名为 DolphinScheduler，简称 "DS"，中文名“小海豚调度”（海豚聪明、人性化，有左右脑可互相换班，终生不用睡觉）。起这个名字就是希望 DolphinScheduler 像它的名字一样，成为一个“开箱即用”的灵活易用的调度系统。

特点：数据处理，可视化DAG，分布式，可扩展

一个分布式易扩展的可视化 DAG 工作流任务调度系统。致力于解决数据处理流程中错综复杂的依赖关系，使调度系统在数据处理流程中 开箱即用。

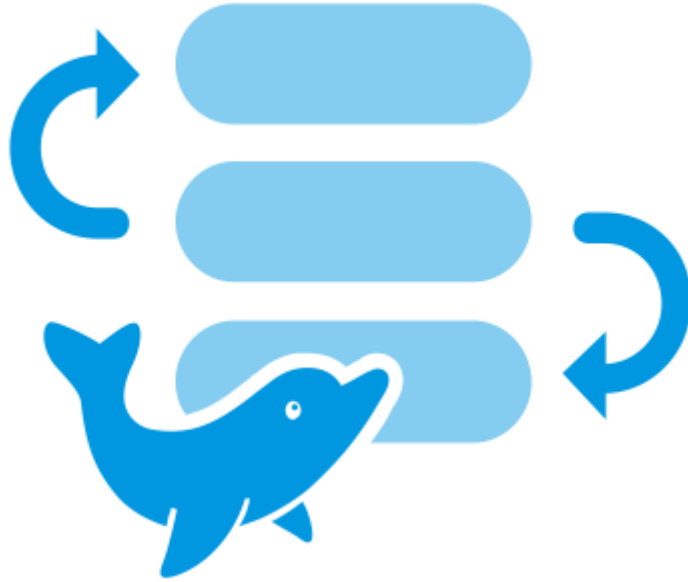
官网：

<https://dolphinscheduler.apache.org/en-us/index.html>

https://github.com/apache/incubator-dolphinscheduler/blob/dev/README_zh_CN.md

试用地址：https://dolphinscheduler.apache.org/en-us/docs/1.3.1/user_doc/quick-start.html

Dolphin Scheduler Logo：



1.2. 和其他调度系统对比

常见的各种调度引擎的特性对比：

	EasyScheduler	Azkaban	Airflow
稳定性			
单点故障	去中心化的多Master和多Worker	是 单个Web和调度程序组合节点	是 单一调度程序
HA额外要求	不需要(本身就支持HA)	DB	Celery / Dask / Mesos + Load Balancer + DB
过载处理	任务队列机制，单个机器上可调度的任务数量可以灵活配置，当任务过多时会缓存在任务队列中，不会造成机器卡死	任务太多时会卡死服务器	任务太多时会卡死服务器
易用性			
DAG监控界面	任务状态、任务类型、重试次数、任务运行机器、可视化变量等关键信息一目了然	只能看到任务状态	不能直观区分任务类型
可视化流程定义	是所有流程定义操作都是可视化的，通过拖拽任务来绘制DAG,配置数据源及资源。同时对于第三方系统，提供api方式的操作。	否 通过自定义DSL绘制DAG并打包上传	否 通过python代码来绘制DAG，使用不便，特别是对不会写代码的业务人员基本无法使用。
快速部署	一键部署	集群化部署复杂	集群化部署复杂
功能			
是否能暂停和恢复	支持暂停，恢复操作	否 只能先将工作流杀死再重新运行	否 只能先将工作流杀死再重新运行
是否支持多租户	支持 easyscheduler上的用户可以通过租户和hadoop用户实现多对一或一对一的映射关系，这对大数据作业的调度是非常重要的。	否	否
任务类型	支持传统的shell任务，同时支持大数据平台任务调度：MR、Spark、SQL(mysql、postgresql、hive、sparksql)、Python、Procedure、Sub_Process	shell、gobblin、hadoopJava、java、hive、pig、spark、hdfsToTeradata、teradataToHdfs	BashOperator、DummyOperator、MySQLOperator、HiveOperator、EmailOperator、HTTPOperator、
契合度	支持大数据作业spark,hive,mr的调度，同时由于支持多租户，与大数据业务更加契合	由于不支持多租户，在大数据平台业务使用不够灵活	由于不支持多租户，在大数据平台业务使用不够灵活
扩展性			
是否支持自定义任务类型	是	是	是
是否支持集群扩展	是 调度器使用分布式调度，整体的调度能力会随便集群的规模线性增长，Master和Worker支持动态上下线	是，但是复杂 Executor水平扩展	是，但是复杂 Executor水平扩展

1.3. 为什么要用DS呢？

Why DolphinScheduler



High Reliability

Decentralized multi-master and multi-worker, HA is supported by itself, overload processing



User-Friendly

All process definition operations are visualized, Visualization process defines key information at a glance, One-click deployment



Rich Scenarios

Support pause, recover operation. Support multi-tenant. Support more task types e.g., spark, hive, mr, shell, python, sub_process



High Expansibility

Support custom task types, Distributed scheduling, and the overall scheduling capability will increase linearly with the scale of the cluster

丰富的特性:

- 以 DAG 图的方式将 Task 按照任务的依赖关系关联起来, 可实时可视化监控任务的运行状态
- 支持丰富的任务类型: Shell、MapReduce、Spark、SQL(mysql、postgresql、hive、sparksqll), Python, Sub_Process、Procedure 等
- 支持工作流定时调度、依赖调度、手动调度、手动暂停/停止/恢复, 同时支持失败重试/告警、从指定节点恢复失败、Kill 任务等操作
- 支持工作流优先级、任务优先级及任务的故障转移及任务超时告警/失败
- 支持工作流全局参数及节点自定义参数设置
- 支持资源文件的在线上传/下载, 管理等, 支持在线文件创建、编辑
- 支持任务日志在线查看及滚动、在线下载日志等
- 实现集群HA, 通过Zookeeper实现Master集群和Worker集群去中心化
- 支持对Master/Worker cpu load, memory, cpu在线查看
- 支持工作流运行历史树形/甘特图展示、支持任务状态统计、流程状态统计
- 支持补数
- 支持多租户
- 支持国际化

1.4. 谁在用DS呢?

Who Is Using DolphinScheduler

Providing your info on Wanted: who's using DolphinScheduler to help improving DolphinScheduler better.



1.5. 初次见面

EasyScheduler

[Home](#)
[Project](#)
[Resources](#)
[Datasource](#)
[Monitor](#)

English apple

Project Home

Process

Process definition

Process Instance

Task Instance

Process definition

Create process

Please enter keyword

Q

	#	Process Name	State	Create Time	Update Time	Description	Timing state	Operation
<input type="checkbox"/>	1	test_j	online	2019-03-14 09:57:20	2019-03-14 19:42:26		offline	
<input type="checkbox"/>	2	DAG	offline	2019-03-15 16:42:30	2019-07-17 18:30:20		-	
<input type="checkbox"/>	3	test	online	2019-03-01 11:28:43	2019-05-12 23:40:56		-	
<input type="checkbox"/>	4	test_h	offline	2019-02-25 16:41:59	2019-03-14 18:31:23		-	
<input type="checkbox"/>	5	newTest	offline	2019-02-23 18:32:10	2019-03-13 23:04:50	new	-	
<input type="checkbox"/>	6	test444	offline	2019-01-25 15:53:38	2019-01-25 18:19:23		-	
<input type="checkbox"/>	7	test_1	offline	2019-01-25 10:04:51	2019-01-25 10:10:10		-	
<input type="checkbox"/>	8	sub5	offline	2018-12-13 11:29:04	2019-01-24 18:28:03		-	
<input type="checkbox"/>	9	test111	offline	2018-12-28 17:43:12	2019-01-22 17:53:55		-	
<input type="checkbox"/>	10	test2	offline	2019-01-21 20:29:30	2019-01-22 09:49:57		-	

<

1

2

3

4

5

6

>

Go to

EasyScheduler

[Home](#)
[Project](#)
[Resources](#)
[Datasource](#)
[Monitor](#)

English apple

Project Home

Process

Process definition

Process Instance

Task Instance

Process definition

Toolbar

Current node settings

Node name

shell_task2

Run flag

Normal

Prohibition execution

Description

Please enter description

Task priority

MEDIUM

Worker group

Default

Number of failed retries

0

(Times)

Failed retry interval

1

(Minute)

Timeout alarm

Script

echo 1

Resources

Please select resources

Custom Parameters

Cancel

Confirm add

EasyScheduler

[Home](#)
[Project](#)
[Resources](#)
[Datasource](#)
[Monitor](#)

English apple

Project Home

Process

Process definition

Process Instance

Task Instance

Process definition

TreeView

Node Type

SHELL

SUB_PROCESS

PROCEDURE

SQL

SPARK

MR

PYTHON

DEPENDENT

Task Status

Submitted successfully

Executing

Ready to pause

Pause

Ready to stop

Stop

Failed

Success

Need fault tolerance

Kill

Waiting for thread

Waiting for dependence

25

EasyScheduler

项目首页

项目管理

资源中心

数据源中心

项目首页

工作流

工作流定义

工作流实例

任务实例

树形图

节点类型 ● SHELL ● SUB_PROCESS ● PROCEDURE ● SQL ● SPARK ● MR ● PYTHON

任务状态 ■ 提交成功 ■ 正在执行 ■ 准备暂停 ■ 暂停 ■ 准备停止 ■ 停止 ■ 失败 ■ 成功 ■ 需要容错 ■ kill ■ 等待线程 ■ 等待依赖

如果想自己体验一下，官方有一个试用账号。<http://106.75.43.194:8888/view/login/index.html>